Behavioral/Cognitive

Selective Enhancement of Object Representations through Multisensory Integration

David A. Tovar,[1,2] Micah M. Murray,[3,4,5,6] and Mark T. Wallace[1,2,6,7,8,9]

[1]School of Medicine, Vanderbilt University, Nashville, Tennessee 37240, [2]Vanderbilt Brain Institute, Vanderbilt University, Nashville, Tennessee 37240, [3]The Laboratory for Investigative Neurophysiology (The LINE), Department of Radiology, Lausanne University Hospital and University of Lausanne (CHUV-UNIL), 1011 Lausanne, Switzerland, [4]Sensory, Cognitive and Perceptual Neuroscience Section, Center for Biomedical Imaging (CIBM) of Lausanne and Geneva, 1015 Lausanne, Switzerland, [5]Department of Ophthalmology, Fondation Asile des aveugles and University of Lausanne, 1002 Lausanne, Switzerland, [6]Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, Tennessee 37240, [7]Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240, [8]Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, Tennessee 37240, and [9]Department of Pharmacology, Vanderbilt University, Nashville, Tennessee 37240

Objects are the fundamental building blocks of how we create a representation of the external world. One major distinction among objects is between those that are animate versus those that are inanimate. In addition, many objects are specified by more than a single sense, yet the nature by which multisensory objects are represented by the brain remains poorly understood. Using representational similarity analysis of male and female human EEG signals, we show enhanced encoding of audiovisual objects when compared with their corresponding visual and auditory objects. Surprisingly, we discovered that the often-found processing advantages for animate objects were not evident under multisensory conditions. This was due to a greater neural enhancement of inanimate objects—which are more weakly encoded under unisensory conditions. Further analysis showed that the selective enhancement of inanimate audiovisual objects corresponded with an increase in shared representations across brain areas, suggesting that the enhancement was mediated by multisensory integration. Moreover, a distance-to-bound analysis provided critical links between neural findings and behavior. Improvements in neural decoding at the individual exemplar level for audiovisual inanimate objects predicted reaction time differences between multisensory and unisensory presentations during a Go/No-Go animate categorization task. Links between neural activity and behavioral measures were most evident at intervals of 100–200 ms and 350–500 ms after stimulus presentation, corresponding to time periods associated with sensory evidence accumulation and decision-making, respectively. Collectively, these findings provide key insights into a fundamental process the brain uses to maximize the information it captures across sensory systems to perform object recognition.

Key words: **decoding; EEG; multisensory integration; object recognition; representational similarity analysis**

Significance Statement

Our world is filled with ever-changing sensory information that we are able to seamlessly transform into a coherent and meaningful perceptual experience. We accomplish this feat by combining different stimulus features into objects. However, despite the fact that these features span multiple senses, little is known about how the brain combines the various forms of sensory information into object representations. Here, we used EEG and machine learning to study how the brain processes auditory, visual, and audiovisual objects. Surprisingly, we found that nonliving (i.e., inanimate) objects, which are more difficult to process with one sense alone, benefited the most from engaging multiple senses.

Received Aug. 27, 2019; revised Apr. 17, 2020; accepted May 21, 2020.

Author contributions: D.A.T., M.M.M., and M.T.W. designed research; M.M.M. performed research; D.A.T. analyzed data; D.A.T., M.M.M., and M.T.W. wrote the paper.

The authors declare no competing financial interests.

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (Grant T32-GM-007347). M.M.M. is supported by the Swiss National Science Foundation (Grant 169206), the Fondation Asile des aveugles (Grant 232933), and a grantor advised by Carigest SA. We are grateful to Céline Cappe for technical assistance during data acquisition and Tijl Grootswagers for helpful comments on earlier drafts of the manuscript.

Correspondence should be addressed to David A. Tovar at david.tovar@vanderbilt.edu.

https://doi.org/10.1523/JNEUROSCI.2139-19.2020

Introduction

The brain is constantly bombarded with sensory information, much of which is combined to form building blocks of our perception representation of the external world. Previous multisensory literature has shown that the brain tends to optimally combine sensory information when the information between senses is equally reliable (Ernst and Banks, 2002). Furthermore, prior work has shown that the maximum gains from multisensory integration are seen when responses to the individual senses are weak (Stein and Meredith, 1993; Wallace et al., 2004). In large measure, these studies have focused on manipulating stimulus reliability and effectiveness through changing low-level stimulus features, such as introducing differing levels of noise, to gauge the effects on multisensory integration. However, emerging literature in vision and audition suggests that higher-level semantic features, such as the binding of stimulus elements into objects, may also play a key role in dictating reliability and effectiveness (Cappe et al., 2012; Ritchie et al., 2015). Given that many objects are specified through their multisensory features, an open question is how might differences in object categorization lead to differences in perceptual gains from multisensory integration.

One of the major categorical distinctions between objects is animacy. In vision, animate objects offer substantial processing and perceptual advantages over inanimate objects, including being categorized faster, more consciously perceived, and found faster in search tasks (New et al., 2007; Jackson and Calvillo, 2013; Carlson et al., 2014; Ritchie et al., 2015; Lindh et al., 2019). Auditory studies have similarly found faster categorization times for animate objects (Yuval-Greenberg and Deouell, 2009; Vogler and Titchener, 2011). This difference may be a remnant of an evolutionary need to rapidly recognize and process living stimuli that could pose threats or be sources of sustenance (Laws, 2000). Furthermore, many inanimate objects such as cars, trains, and cellphones have not existed long enough for there to be specialized brain areas to represent them. In contrast, a number of specialized areas exist for the processing of categories of animacy, such as faces in the fusiform face area, bodies in the extrastriate body area, and voices in the temporal voice areas (Kanwisher et al., 1997; Belin et al., 2000; Downing et al., 2001; De Lucia et al., 2010).

To study how perceptual differences in visual and auditory categories influence their subsequent integration as audiovisual objects, it is critical to quantify neural encoding differences between objects. Representational similarity analysis (RSA; Kriegeskorte et al., 2008a) constructs a representational space quantifying relationships between stimuli with representational distance indicating the difference in their neural signatures. A greater distance in representational space signifies more distinct neural signals between stimuli, while shorter distances signify less distinct neural signals. Studies using RSA have shown that visual and auditory objects have a clear encoding distinction between animate and inanimate categories (Kriegeskorte et al., 2008b; Giordano et al., 2013; Cichy et al., 2014), while also showing that representational space can contract if stimuli are degraded (Grootswagers et al., 2017b) or expand in cases of increased attention (Nastase et al., 2017). Although RSA has been increasingly used to study object representations, it has not been fully leveraged to examine objects as they are often represented in naturalistic setting as multisensory entities.

In this study, we presented subjects with auditory, visual, and semantically congruent audiovisual animate and inanimate objects while we recorded high-density EEG. Our overarching hypothesis was that greater behavioral benefits would be seen for objects specified in a multisensory manner and that these gains would be accompanied by an expansion in representational space as measured using RSA. A secondary hypothesis was that greater benefits would be observed for inanimate objects, given evidence that multisensory integration benefits are greatest for weakly effective stimuli (Stein and Meredith, 1993; Ernst and Banks, 2002; Wallace et al., 2004).

Materials and Methods

Participants. The experiment included 14 adults (9 males) with a mean age of 27 ± 4.2 years. All subjects had normal or corrected-to-normal vision and reported normal hearing. The study was conducted in accordance with the Declaration of Helsinki, and all subjects provided their informed consent to participate in the study. Each participant was compensated financially for their participation. The experimental procedures were approved by the Ethics Committee of the Vaudois University Hospital Center and University of Lausanne. Behavioral data for all subjects were used. However, EEG data for one subject was removed from further decoding analysis due to poor signal quality in the evoked potential response.

Stimuli. The experiment took place in a sound-attenuated chamber (Whisper Room), where subjects were seated centrally in front of a 20 inch computer monitor (LP2065, HP) and located ∼140 cm away from them (visual angle of objects, ∼4°). The auditory stimuli were presented over insert earphones (model ER4S, Etymotic Research), and the volume was adjusted to a comfortable level (∼62 dB). The stimuli were presented and controlled by E-Prime 2.0, and all behavioral data were recorded in conjunction with a serial response box (Psychology Software Tools; https://www.pstnet.com/). The auditory stimuli included 48 animate and 48 inanimate sounds from a library of 500-ms-duration sounds, used in previous studies and have been evaluated in regard to their acoustics and psychoacoustics as well as brain responses as a function of semantic category (Murray et al., 2006; De Lucia et al., 2010; Thelen et al., 2012). The visual stimuli were semantically congruent line drawings that were taken from a standardized set (Snodgrass and Vanderwart, 1980) or obtained from an online library (http://dgl.microsoft.com).

Experiment design. Participants performed 10–13 experimental blocks (median, 10 blocks) of a Go/No-Go task. Each block contained 1 audio, visual, and audiovisual presentation for each of the 96 stimuli exemplars, totaling 288 stimulus presentations per block. For half of the blocks, subjects were instructed to press a button when they perceived an animate object and for the other half when they perceived an inanimate object. Animate and inanimate blocks were randomized for each subject. Auditory, visual, and synchronous audiovisual stimuli were presented for 500 ms, followed by a randomized interstimulus interval ranging from 900 and 1500 ms, and participants had to respond within this 1.4–2 s window. Stimuli modality was randomized for each trial (Fig. 1, schematic). To control for motor confounds, the block instructions alternated between indicating whether the stimuli were animate or inanimate (Grootswagers et al., 2017a). Reaction times (RTs) and accuracy were measured for each response. Participants did not receive feedback during the experiment.

Statistical inference. All statistical inference for behavior and neural data were assessed with Bayes factors (BFs; Jeffreys, 1998; Wetzels et al., 2011) using a JZS (Jeffreys–Zellner–Siow) prior (Rouder et al., 2009), with a scale factor of 0.707. For decoding analysis, chance-level decoding was estimated by randomly shuffling all trial labels for each subject once before classification to construct a null distribution. The probability of the group data assuming the alternative hypothesis relative to the probability of group data assuming chance-level decoding was computed to calculate a Bayes factor at each time point. Bayes factors provide the added advantage over frequentist inference because in addition to rejecting the null hypothesis, they can provide support for the null hypothesis as well as determine whether the data are insensitive, and as a result help avoid overstating the evidence against the null hypothesis (Edwards et al., 1963; Berger and Delampady, 1987; Sellke et al., 2001; Johnson, 2013). The theoretical differences underlying Bayesian and frequentist analyses have spurred debate on whether and how Bayes factors should be corrected for multiple comparisons (Berry and Hochberg, 1999), since they intrinsically already reduce type I errors (Gelman and Tuerlinckx, 2000; Gelman et al., 2012; Johnson, 2013). In this study, we report Bayes factors without additional multiple-comparison correction, but provide Bayes factors with varying levels of evidence, consistent with recent EEG decoding studies (Grootswagers et al., 2019; Robinson et al., 2019). Using Jeffreys' scheme, Bayes factors >3 and >10 indicate substantial and strong evidence for the alternative hypothesis, respectively, anything between 3 and 1/3 indicates insufficient evidence, and Bayes factors less than 1/3 and 1/10 indicate substantial and strong evidence for the null hypothesis (Jeffreys, 1998; Jarosz and Wiley, 2014). We further compared Bayes factors with a cluster-based sign permutation test (Maris and Oostenveld, 2007) and found Bayes factors to be more conservative. Therefore, we report only Bayes factors in the Results.

EEG acquisition and preprocessing. A continuous EEG was acquired from 160 scalp electrodes (sampling rate, 1024 Hz) using a Biosemi ActiveTwo System. Data preprocessing was performed offline using the Fieldtrip toolbox (Oostenveld et al., 2011) in MATLAB. Data were filtered using a Butterworth IIR filter with 1 Hz high pass, 60 Hz low pass, and notch at 50 Hz. All channels were rereferenced to an average reference. Epochs were created for each stimulus presentation ranging from −100 to 600 ms relative to stimulus onset. Each epoch was baseline corrected using the prestimulus period.

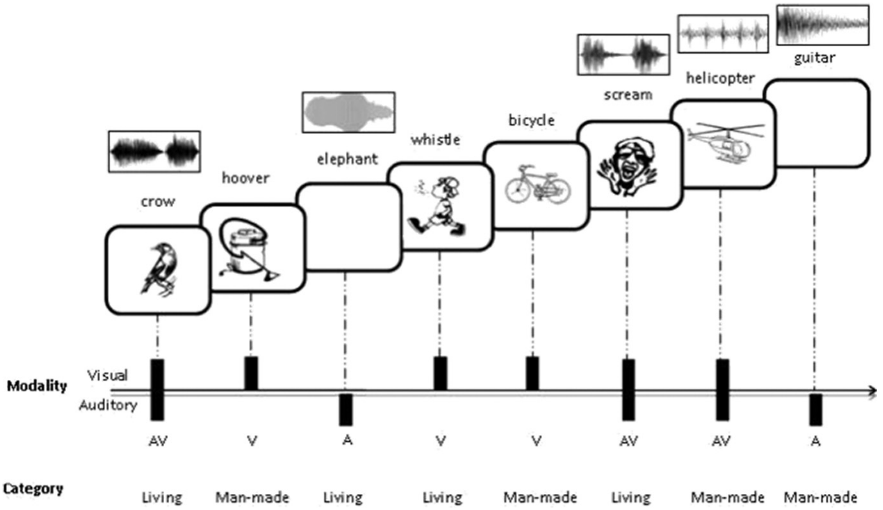

Figure 1. Experiment schematic. A Go/No-Go discrimination task of animate and inanimate objects. The responses were counterbalanced such that the number of responses for animate and inanimate objects was equivalent. The stimuli consisted of 96 visual line drawings and 96 environmental sounds of common animate and inanimate objects, as well as semantically congruent pairings of these objects. The sounds of animate objects were nonverbal vocalizations. The stimulus duration was 500 ms with a variable interstimulus interval of 900–1500 ms.

Representational similarity analysis. Following data preprocessing, we used CoSMoMVPA (Oosterhof et al., 2016) and custom scripts to perform cross-validated RSA. We used a linear discriminate classifier after default regularization (0.01) with fourfold, leave-one-fold-out cross-validation, for all exemplar pair combinations across audio, visual, and audiovisual stimuli presentations. In this procedure, trials are randomly assigned to one of four subsets of data. Three of the four subsets (75% of the data) are then pooled together to train the classifier, and then decoding accuracy is tested on the remaining subset (25% of the data). This procedure is repeated a total of four times, such that each of the subsets is tested at least once. Decoding results are reported in the percentage correct of classifications at each time point for each exemplar pair in the time series [−100, 600 ms]. This analysis was conducted independently to build representational dissimilarity matrices (RDMs) for each subject and modality over 1 ms increments. The RDMs were then separated into animate exemplar pairwise comparisons, inanimate exemplar pairwise comparisons, and pairwise comparisons between categories. Using these comparison groupings, mean decoding accuracies were then calculated for each modality and subject. Significant above-chance accuracies were assessed against a randomized trial shuffle control using Bayes factors.

Representational connectivity analysis. To characterize connectivity changes for different modalities and object categories, we used a combination of a searchlight analysis and representational connectivity analysis (Kriegeskorte et al., 2008b). Because of this analysis being computationally intense, data were downsampled to 100 Hz. Electrode-specific RDMs, using the same procedure described for the RSA, were built by using a moving searchlight that included the electrode of interest and every immediate adjacent electrode. Depending on the location of the electrodes, the RDMs can potentially be more descriptive of lower-level properties of the stimuli or contain higher-level object category information. Importantly, the analysis is not designed to distinguish between any particular stimulus dimension, such as animacy, but rather used to calculate the local representational geometry present at those electrodes. Electrode-specific RDMs were then correlated to each other in pairwise fashion for each electrode combination using a Spearman correlation to form a matrix of RDM correlations between electrodes. We then averaged the Spearman correlations from across all electrode comparisons to compute a mean connectivity measure. If the representational geometry is distributed across several electrodes, then the expectation is that this value would increase, and, if it is unique to a particular electrode, this value would decrease. This analysis was performed for visual, auditory, and audiovisual presentations. Additionally, to compare the audiovisual response to the visual and auditory response more directly, we also summed evoked responses for auditory and visual presentations for each specific exemplar and performed RCA on these trials.

Note that, in this calculation, the searchlight will change sizes depending on the chosen electrode and searchlights will overlap for electrodes leading to a nonzero baseline level of connectivity in neighboring RDMs, regardless of the evoked responses to stimulus presentations. Therefore, we repeated the analysis above, but shuffled all of the exemplar labels when calculating the RDMs to create a shuffled control. All connectivity measurements were compared with their respective shuffled labels control. This procedure was done for all exemplars as well as within the animate and inanimate category along the time series [−100 600 ms] to compute time-resolved representational connectivity measures.

Distance-to-bound analysis. To link neural representational space back to individual exemplar categorization times, we used a distance-to-bound analysis (for review, see Ritchie and Carlson, 2016). Similar to RSA, this analysis represents individual exemplars as points in representational space. A decision boundary for animacy is then fitted using a linear discriminant analysis classifier to the representational space, defining an optimal decision boundary that separates animate and inanimate exemplars. The distance to the decision boundary is determined for each exemplar and subsequently pooled and averaged across subjects to calculate average exemplar distance across subjects for each time point in the time series [−100 600 ms]. Next, the median exemplar reaction time, pooled across subjects, is calculated for each exemplar. We then performed a time-varying Spearman correlation between mean exemplar distance and median exemplar reaction time for each modality using a fixed-effects analysis to reduce noise and improve statistical power. The distance-to-bound analysis was performed across all electrodes as well as on an electrode by electrode basis using a moving searchlight.

Model fitting. To account for low-level visual features in our visual and auditory stimuli, we constructed model RDMs and calculated their correlations to electrode-specific RDMs and the neural RDM from all electrodes. The low-level feature auditory RDM was constructed using a Welch's power spectral density (PSD) estimate for each of the 96 sounds. The resulting stimulus PSD was then organized into vectors, and pairwise nonparametric Spearman distance measurements were calculated for all exemplar pair combinations to form a model RDM. We then calculated the Spearman correlations between the PSD model RDM and the modality-specific neural RDMs at each time point. An identical procedure was followed for the visual images, but instead of using PSD, image contrast was used. Note that since the images were black and white Snodgrass images, the contrast values will be equivalent to the image intensity values. In addition to these low-level feature models, we also constructed an abstract animacy category model. The animacy

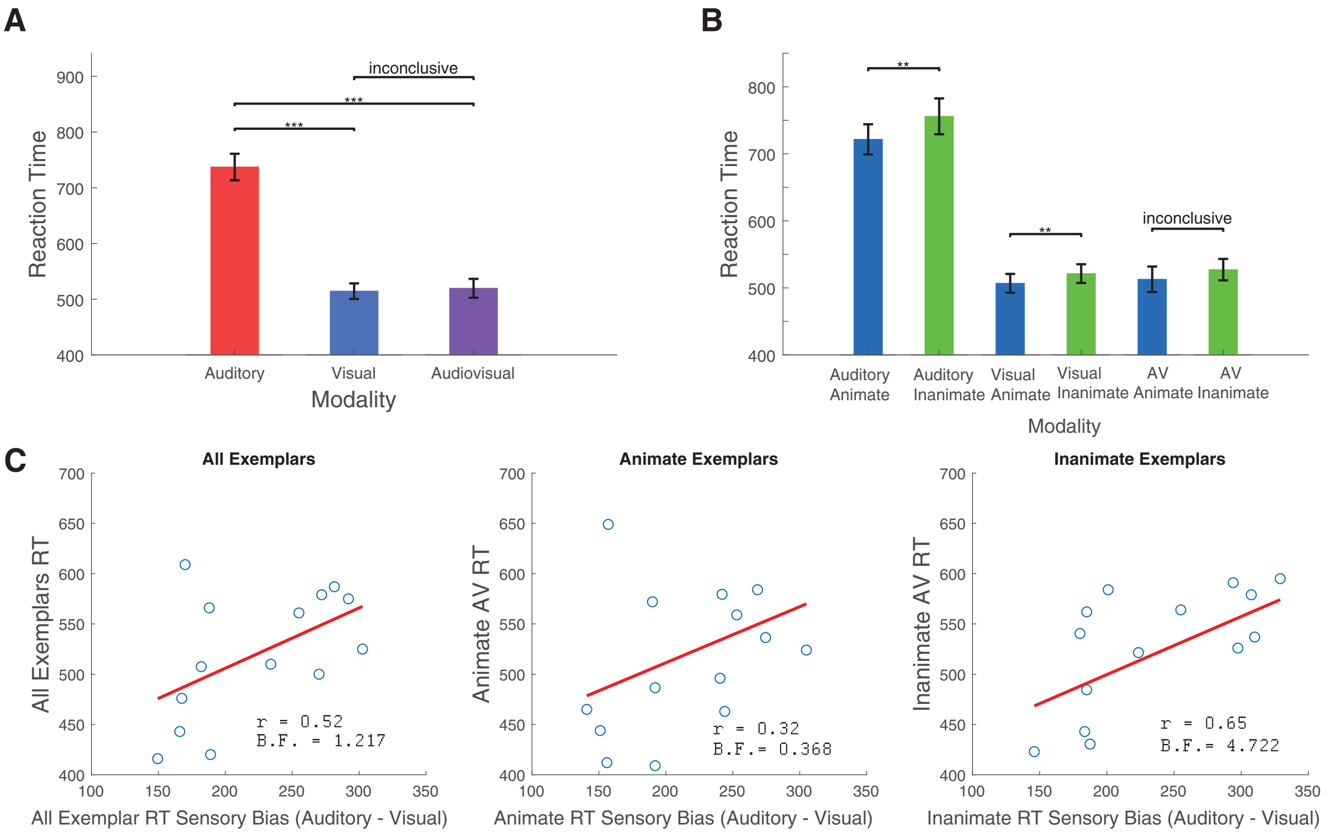

Figure 2. Behavior: advantage for animate objects for unisensory presentations but not audiovisual presentations. ***A***, ***B***, RT results for each modality (***A***) and broken down by animacy (***B***). Bayes factors for substantial evidence (*BF $>$ 3), strong evidence (**BF $>$ 10), and very strong evidence (***BF $>$ 30) above comparisons. ***C***, Subject sensory bias and audiovisual RT Pearson correlations across subjects for all exemplars, only animate exemplars, and only inanimate exemplars. Sensory bias is only significantly correlated to audiovisual RT for inanimate exemplars (BF $>$ 3).

category model was constructed using a 0 to indicate no differences between stimuli pairs for within-animacy category exemplars and a 1 to indicate complete dissimilarity for between-category exemplars. This model was then also tested across modality-specific neural RDMs.

Results

Behavior: advantage for animate objects on unisensory but not multisensory (i.e., audiovisual) presentations

Subjects were shown 48 animate and 48 inanimate auditory, visual, and audiovisual objects while they performed a Go/No-Go categorization task, as shown in Figure 1. Subjects performed near ceiling on the categorization task for objects presented in both visual (animate, 98%; inanimate, 98%) and audiovisual (animate, 98%; inanimate, 99%) contexts, and were less accurate for auditory presentations (animate, 86%; inanimate, 87%). A two-way repeated-measures ANOVA for accuracy revealed a main effect for modality ($F_{(2,26)} = 27.14$, $p = 0.00$), but no main effect for animacy ($F_{(1,26)} = 0.64$, $p = 0.44$).

When examining RTs, a two-way repeated-measures ANOVA revealed main effects for modality ($F_{(2,26)} = 238.18$, $p = 0.00$) and animacy ($F_{(1,26)} = 10.39$, $p = 0.01$), as well as an interaction effect ($F_{(2,26)} = 3.68$, $p = 0.04$). We then performed *post hoc* tests across sensory modalities and categories, as shown in Figure 2. Figure 2*A* shows median RTs for the Go/No-Go task across participants for the three sensory conditions. Using Bayes factors to compare median RTs across subjects, we found very strong evidence (BF $>$ 30) that the auditory condition was slower than the visual and audiovisual conditions. Next, behavior was split by animate and inanimate categories to investigate the effects of animacy on RTs. Figure 2*B* shows that there was strong evidence (BF $>$ 10) for faster RTs for animate objects compared with inanimate objects when presented in either the auditory or visual modalities, consistent with the results from previous studies (Murray et al., 2006; Yuval-Greenberg and Deouell, 2009; Vogler and Titchener, 2011; Carlson et al., 2014). However, there was inconclusive evidence (BF = 0.75) for the audiovisual condition.

To further investigate this surprising lack of a difference in audiovisual performance, we created an index of sensory bias for each participant, operationalized as the difference in reaction times to the auditory and visual stimuli, and correlated this bias score to audiovisual RTs on a subject-by-subject basis using a Pearson correlation. Figure 2*C* shows that the only significant correlation between sensory bias and audiovisual RTs was for inanimate objects. The positive correlation indicates that subjects whose RTs for visual and auditory stimuli were more similar had faster multisensory RTs. Note, that these correlations included all subjects, since there were no outliers for sensory bias or audiovisual reaction times.

Representational similarity analysis: the influence of sensory modality on between and within animacy category decoding

To investigate the neural correlates of the behavioral differences noted across conditions, we used RSA (Fig. 3*A*–*C*). Specifically, we built RDMs for each subject and modality over 1 ms intervals using linear discriminant analysis for each exemplar pair. From each RDM, we explored the effect of sensory modality on the distinction between animate and inanimate exemplars by

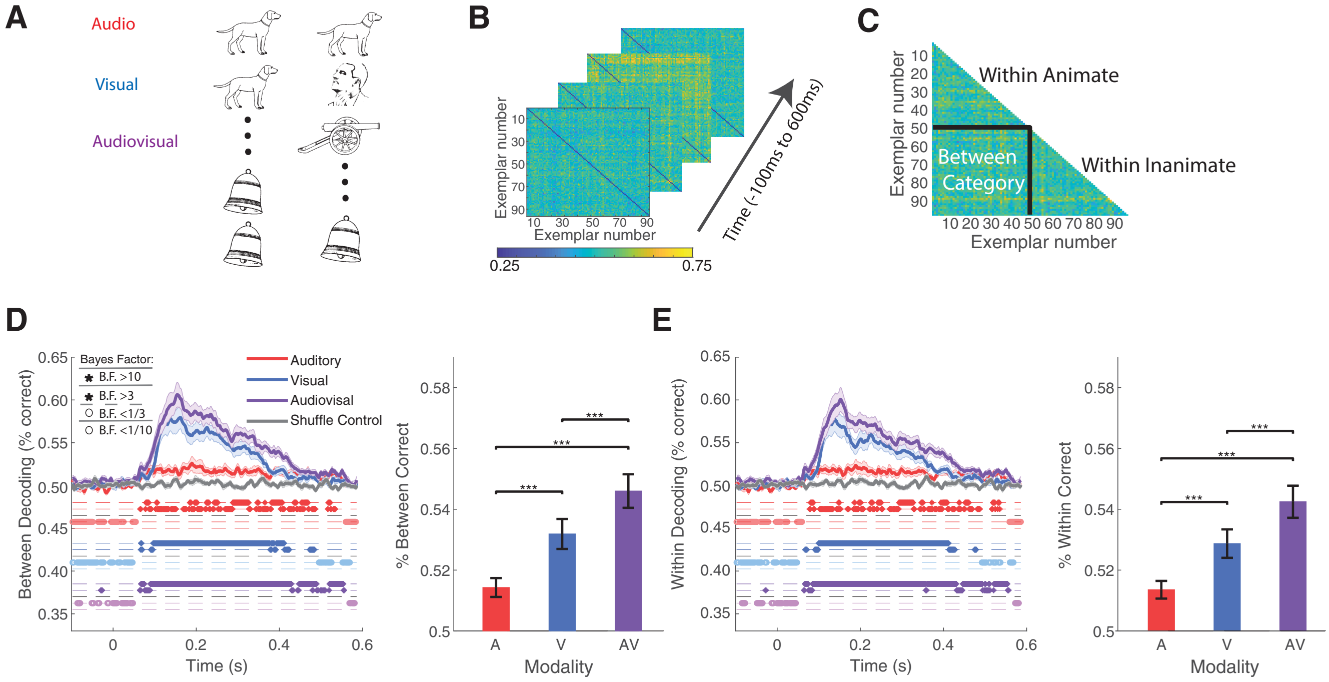

Figure 3. Representational similarity analysis: sensory modality influences between-animacy category and within-animacy category decoding. ***A***, RSA schematic for pairwise decoding. Linear discriminate analysis with a fourfold leave-one-fold-out cross-validation was used for all exemplar pair combinations. ***B***, Dissimilarity matrices for each of the modalities was built across time in 1 ms increments from pairwise exemplar classifications. ***C***, Mean between-category and within-category exemplar decoding accuracies were averaged across exemplars at each time point. ***D***, ***E***, Resulting time series and summary bar plots for between (***D***) and within (***E***) categories for each of the modalities. Shaded area around lines indicates standard error across subjects. Asterisks indicate thresholded Bayes factors for alternative and null hypotheses (see inset). Mean decoding across time (50–500 ms) for each modality with Bayes factors for substantial evidence (*BF $>$ 3), strong evidence (**BF $>$ 10), and very strong evidence (***BF $>$ 30) above comparisons.

calculating the mean pairwise decoding for between-category pairs (e.g., dog vs bell, dog vs cannon). As can be seen in Figure 3*D*, before stimulus onset, decoding is close to the shuffled label control at chance level (i.e., 50%), because the classifier does not have any meaningful neural data that will distinguish between-category pairs. However, shortly after stimulus onset, decoding performance becomes significantly above the shuffled label control (BF $>$ 3) across all three modalities. The latency of the onset of these decoding differences, defined as at least 20 ms of sustained significant decoding (Carlson et al., 2013), was 183 ms for auditory, 91 ms for visual, and 65 ms for audiovisual stimulus conditions. Visual and audiovisual decoding peaked at 162 and 154 ms, respectively, with higher absolute peak decoding for audiovisual presentations (61%) compared with visual presentations (58%). Decoding of auditory stimuli was comparatively poorer, peaking at 53% at 190 ms. Note that while there were differences in significant decoding onsets, caution should be taken when comparing decoding onsets across conditions with different maximum decoding peaks (Grootswagers et al., 2017a, their Fig. 14). Collectively, the results of these decoding analyses illustrate the temporal emergence of distinct neural representations for auditory, visual, and audiovisual objects when subjects are performing an animacy/inanimacy categorization.

To statistically compare decoding performance across modalities, we computed the mean decoding for the interval spanning 50–500 ms post-stimulus presentation. When comparing mean decoding values across subjects, audiovisual stimuli were significantly higher when compared with both visual and auditory decoding (BF >30), and visual decoding was higher than auditory decoding. These modality-focused RSA results suggest that the audiovisual presentation of an object creates a more distinct representation between animate and inanimate objects when compared with either of the corresponding unisensory presentations.

We further explored whether audiovisual presentations expanded exemplar distinctions within animacy categories by calculating the mean within category pairwise decoding accuracies (Fig. 3*E*). In this analysis, onset latencies for significant decoding for auditory, visual, and audiovisual stimuli were 184, 91, and 79 ms, respectively. The corresponding peak decoding latencies were 189, 139, and 152 ms. The modality-specific comparisons for within-category decoding mirrored those seen for between-category decoding, with higher audiovisual decoding when compared with visual and auditory decoding, and higher visual decoding than auditory decoding (BF $>$ 30). A comparison of between-category decoding and within-category decoding demonstrated higher between-category decoding for auditory, visual, and audiovisual stimulus presentations (BF $>$ 3) during the stimulus period (50–500 ms). In sum, when compared with unisensory presentations, audiovisual stimulus presentations not only expand the representational space between animacy categories, but also make exemplars within the animacy categories easier for a classifier to distinguish.

Category-specific RSA: audiovisual presentations selectively enhance inanimate object decoding

We further investigated representational space broken down by animacy categories to study the neural underpinnings for the observed reaction time differences between animate and inanimate categorization (Fig. 4). The decoding curves for animate and inanimate exemplars did not differ for auditory conditions (Fig. 4*A*) with evidence for the null hypothesis present throughout the time course. However, this was not the case for visual exemplars, which have higher decoding performance for animate exemplars when compared with inanimate exemplars from 160 to 184 ms and from 220 to 228 ms after stimulus presentation

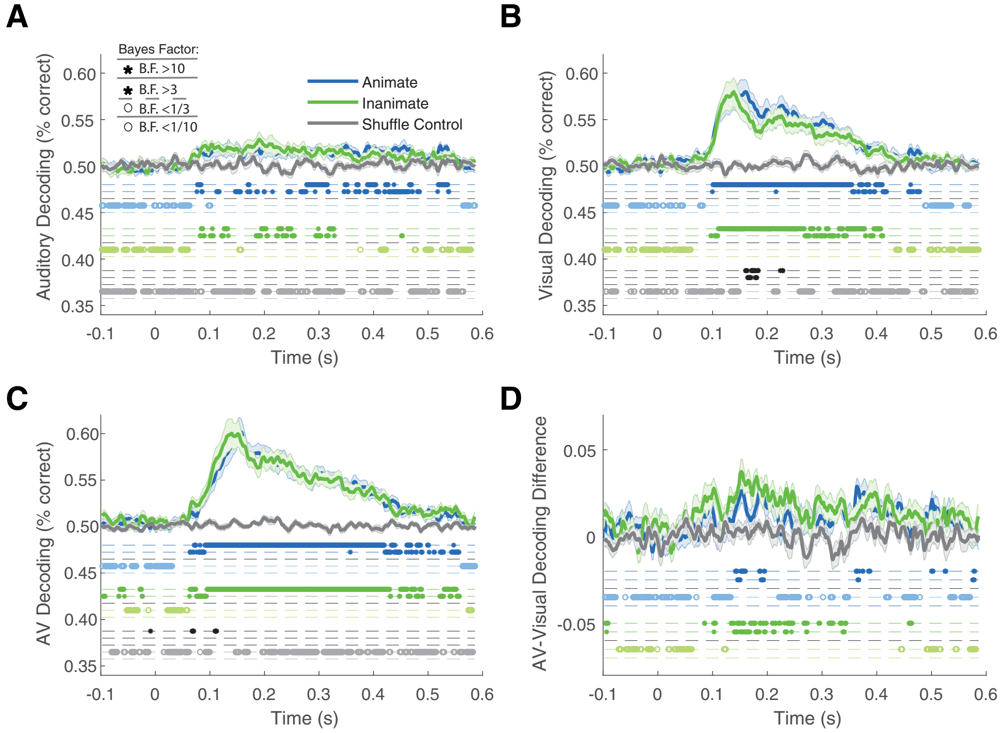

Figure 4. Category-specific RSA: audiovisual presentations selectively enhance inanimate object decoding. ***A–C***, Audio, visual, and audiovisual within-category decoding for animate and inanimate exemplars. Colored asterisks indicate substantial evidence and strong evidence (see inset) compared with the shuffled control, while black asterisks indicate substantial and strong evidence for a difference between animate and inanimate objects. ***D***, The audiovisual-visual within-category decoding difference for animate and inanimate exemplars, with asterisks indicating evidence (see inset) for differences from the shuffle control.

(Fig. 4*B*). Surprisingly, this difference is no longer apparent for audiovisual conditions with in fact a few sporadic timepoints with substantial evidence (BF > 3) that inanimate objects have higher decoding than animate objects.

Since the audiovisual condition had overall higher within-category pairwise decoding than the visual condition (Fig. 3*E*), we additionally wanted to explore whether the lack of an animate and inanimate within-category decoding difference for audiovisual presentations was due to visual inanimate objects incurring a special benefit from audiovisual presentation. Figure 4*D* shows the difference between audiovisual decoding and visual decoding for animate and inanimate exemplars. Notably, the difference is significantly above a shuffle control subtraction of visual and audiovisual presentations for a sustained period of time extending from 137 to 216 ms post-stimulus onset for inanimate objects (BF > 3) but is much sparser for animate objects without a significant sustained difference ever exceeding 20 ms.

Representational connectivity analysis: response patterns between areas in the brain are influenced by modality and object category

Given that different sensory modalities and different object classes have been shown to engage different brain networks (Hillebrandt et al., 2014; Braga et al., 2017), we investigated whether the pairwise decoding differences we found using RSA would also be associated with differences in mean connectivity. To carry out this analysis, we constructed electrode-specific RDMs and performed Spearman correlations across all electrode combinations to calculate a mean representational connectivity measure between electrodes. The mean representational connectivity measure is an index of how similar the representational space is between electrodes. This value is driven by the following two factors: spatial proximity (i.e., neighboring electrodes will have higher connectivity) and representational similarity due to stimulus features. As a control, we performed the analysis on shuffled labels for each of the respective stimulus modalities, which will account for the shared signal due to spatial proximity of neighboring electrodes, but not for the evoked responses to the specific stimuli. The shuffled control served as our comparison for all statistical comparisons.

We found that auditory, visual, and audiovisual presentations all diverged from the shuffled control (BF > 3), beginning at 97, 107, and 78 ms after stimulus presentation, respectively. Averaging across the 50–500 ms stimulus period, we found that audiovisual presentations had more mean connectivity than visual presentations (BF >10) and auditory presentations (BF >30), but there was inconclusive evidence between visual and auditory connectivity (BF = 0.52). In addition, to compare the audiovisual response to the visual and auditory response more directly, we summed the evoked potentials for auditory and visual stimuli for each individual exemplar and then used this summed potential as input to the RCA. We found that the summed unisensory mean connectivity was significantly lower (BF > 30) than the mean audiovisual representational connectivity. These results suggest that shared representations across areas that lead to an increase in the mean connectivity for audiovisual presentations is due to the simultaneous processing of auditory and visual stimuli, and not simply due to visual and auditory signals collectively activating more (or at least a more extensive set of) areas in the brain.

Similar to the RSA findings, we also found that the animate and inanimate category selectively affected connectivity measurements across the different sensory modalities. For auditory objects, connectivity diverged from the shuffled control for animate and inanimate exemplars at 156 ms. Mean connectivity

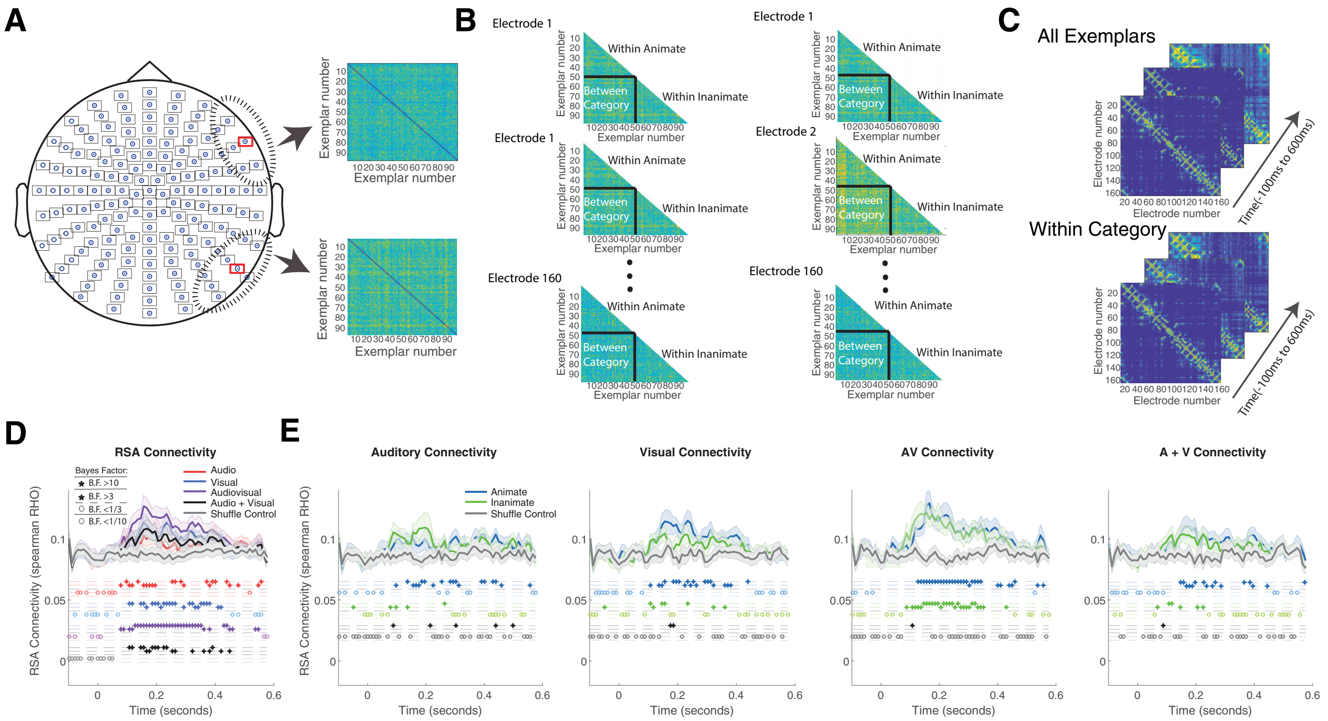

Figure 5. Representational connectivity analysis: response patterns between brain networks are influenced by object category. ***A***, Moving searchlight to create electrode-specific RDMs. The searchlight included the electrode of interest and every immediate surrounding electrode to produce an electrode-specific RDM for each modality. ***B***, Each electrode was correlated in a pairwise fashion using a Spearman correlation. ***C***, This procedure was performed for all exemplars as well as within the animate and inanimate categories along the time course (—100 to 600 ms) to build time-resolved electrode similarity matrices of representational space. The mean value of these matrices is the representational connectivity across all electrodes. ***D***, ***E***, Representational connectivity was measured across modalities and summed unisensory responses (***D***) as well as within the animate and inanimate categories across modalities (***E***). Colored asterisks indicate substantial and strong evidence (see inset) compared with the shuffled control.

over the stimulus period between groups showed substantial evidence for the null hypothesis (BF < 1/3), indicating no animacy difference for representational connectivity in audition. For visual objects, mean connectivity for animate objects and inanimate objects began to diverge from the shuffled control at 137 and 107 ms, respectively. However, visual animate exemplars had a greater mean representational connectivity than inanimate exemplars from 176 to 186 ms and summed over the stimulus period (BF > 3). For audiovisual presentations, inanimate objects diverge from baseline earlier at 107 ms compared with 127 ms for animate objects. In contrast to visual presentations, audiovisual animate and inanimate categories showed inconclusive evidence over the stimulus period (BF = 0.39). Last, for the summed unisensory responses, animate and inanimate objects diverged from the shuffled control at 146 and 107 ms, respectively. Averaged over the stimulus period, there was inconclusive evidence (BF = 0.71) for group differences. In summary, these results build off of the RSA analyses, and suggest that the presentation of objects in an audiovisual manner increase the representational connectivity when compared with when they are presented in a unisensory context, and furthermore that these connectivity measures increase to a greater extent for inanimate exemplars (Fig. 5).

Distance-to-bound analysis: behavior can be predicted by exemplar distance to the decision boundary in representational space

Having found both behavioral and neural differences between the modality of presentation and animacy categories, we next considered whether the two measures were associated with one another. To do this, we computed the distance to the classifier decision boundary for all exemplars and correlated these distances with behavioral performance (i.e., reaction times). A negative correlation would denote that the exemplars that are furthest away from the classifier decision boundary are those that are most rapidly categorized. Indeed, Figure 6*A* shows substantial evidence for a significant negative Spearman correlation (BF > 3) between representational distance and reaction time at several timepoints between 100 and 200 ms post-stimulus onset for both visual and audiovisual presentations, and between 270 and 400 ms post-stimulus onset for all sensory modalities. Below the time course, we show the results from the topographic results from applying the distance-to-bound analysis using a moving searchlight. We found that for visual and audiovisual presentations, occipital and temporal electrodes were most correlated to behavior for the time period spanning 100–200 ms post-stimulus onset. In contrast, frontoparietal electrodes were most correlated with behavior for the interval spanning 270–400 ms post-stimulus onset across all modalities. Figure 6*B* shows the corresponding scatter plot for the highest negative correlations in the 100–200 ms time window for visual and audiovisual presentations. These plots show that for both visual and audiovisual presentations, inanimate objects had slower categorization times than animate objects and were also closer to the decision boundary. Additionally, consistent with our behavioral and RSA results, inanimate exemplars appeared to show a greater shift along the reaction time and representational axes than animate exemplars when comparing between visual and audiovisual scatter plots.

In Figure 6*C*, we quantified this observation by using a Spearman correlation to link the reaction time difference for audiovisual versus visual exemplars with the representational difference for animate and inanimate exemplars. A negative

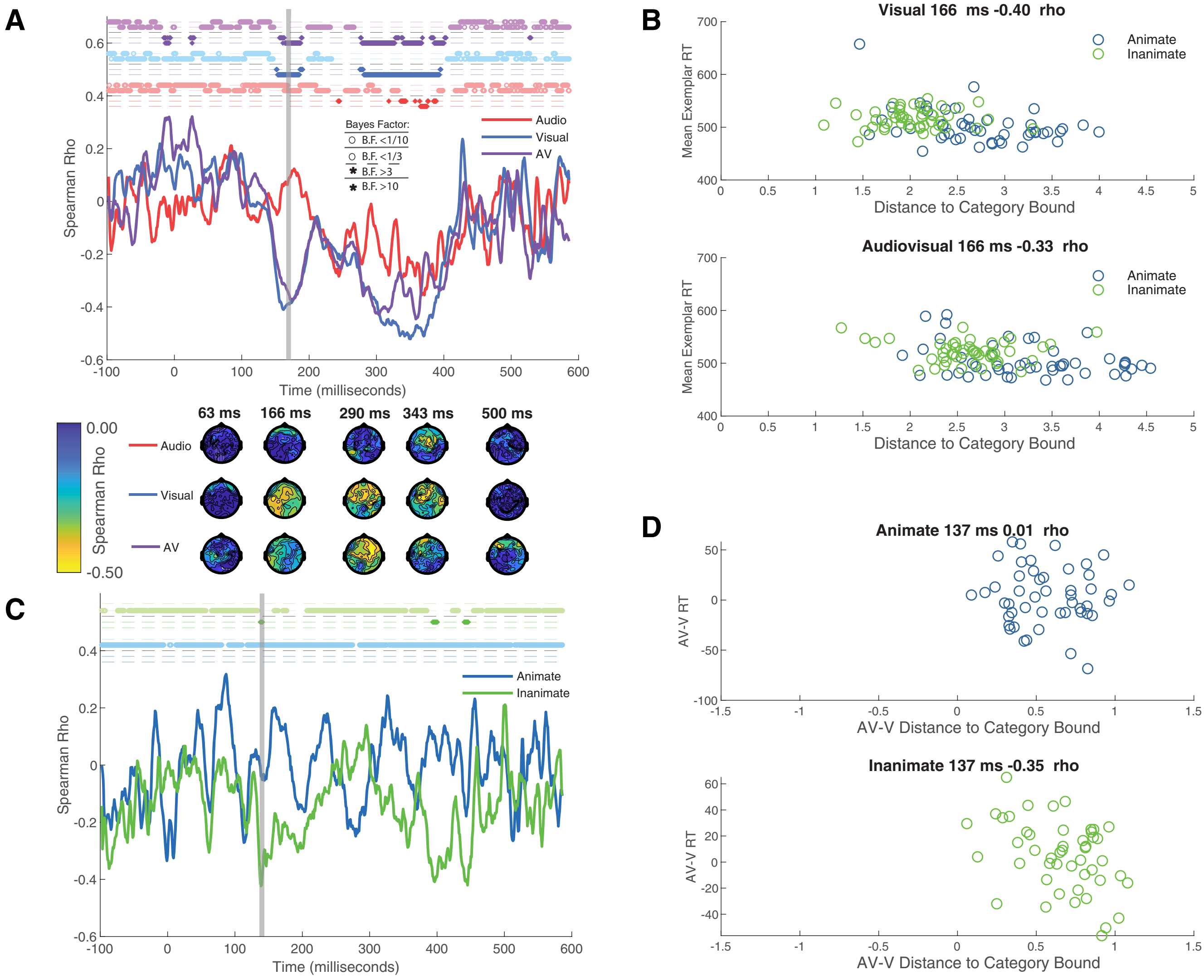

Figure 6. Distance-to-bound analysis: behavior can be predicted by exemplar distance to the decision boundary in representational space. ***A***, Time-varying Spearman correlation between the mean exemplar representational distance from the animacy discriminate bound and the respective average exemplar reaction time for each modality. Asterisks indicate substantial and strong evidence for the alternative hypothesis (BF > 3 and >10) of a nonzero correlation and null hypothesis (BF < 1/3 and <1/10). Below the *x*-axis, results from the topographic results from applying the distance-to-bound analysis using a moving searchlight for select timepoints. ***B***, Scatterplot for mean exemplar visual and audiovisual representational distance and RT at a time point with substantial evidence for the alternative hypothesis for both modalities. ***C***, Time-varying Spearman correlation between mean representational enhancement (AV-V distance) and median reaction time enhancement (AV-V RT), with asterisks indicating evidence for the alternative and null hypothesis. ***D***, Scatterplot for audiovisual representational distance and RT enhancement at a time point with substantial evidence for the alternative hypothesis for inanimate exemplars.

correlation denotes the following: (1) exemplars that were further away from the decisional boundary for audiovisual presentations when compared with visual presentations [positive audiovisual-visual (AV-V) distance value] are also the exemplars that demonstrated either more of an audiovisual RT bias (positive AV-V RT value) or less of a visual bias (negative AV-V RT value); and (2) exemplars that were further away from the decision boundary for visual presentations when compared with audiovisual presentations (negative AV-V distance value) are also the exemplars that demonstrated less of an audiovisual RT bias (positive AV-V RT value) or more of a visual bias (negative AV-V RT value). We found significant timepoints between 100 and 200 ms and 370–450 ms post-stimulus onset supporting the alternative hypothesis (BF > 3) for inanimate exemplars, but only evidence for a null correlation (BF < 1/3) for animate exemplars. If we pool the correlations across the entire stimulus analysis epoch (50–500 ms poststimulus), we find very strong evidence for a negative correlation for inanimate exemplars (BF > 30) but inconclusive evidence for animate exemplars (BF = 2.00). Figure 6*D* shows the corresponding scatterplot with the highest negative correlation in the 100–200 ms window for visual and audiovisual presentations at 137 ms (Fig. 6*B*). Collectively, these results show associations between neural decoding differences, and behavioral performance differences between audiovisual and visual stimulus presentations, but only when these stimuli are inanimate.

Model testing: abstract category models predict neural activity better than low-level feature models

To account for the potential contribution of low-level features to the neural RDMs, we constructed contrast dissimilarity matrices for images and power spectral density dissimilarity matrices for sounds, as shown in Figure 7. The models were correlated using a Spearman correlation to each subject's neural RDM across channels and neighborhoods of electrodes using a moving searchlight to build topographic maps. Along the time series, we

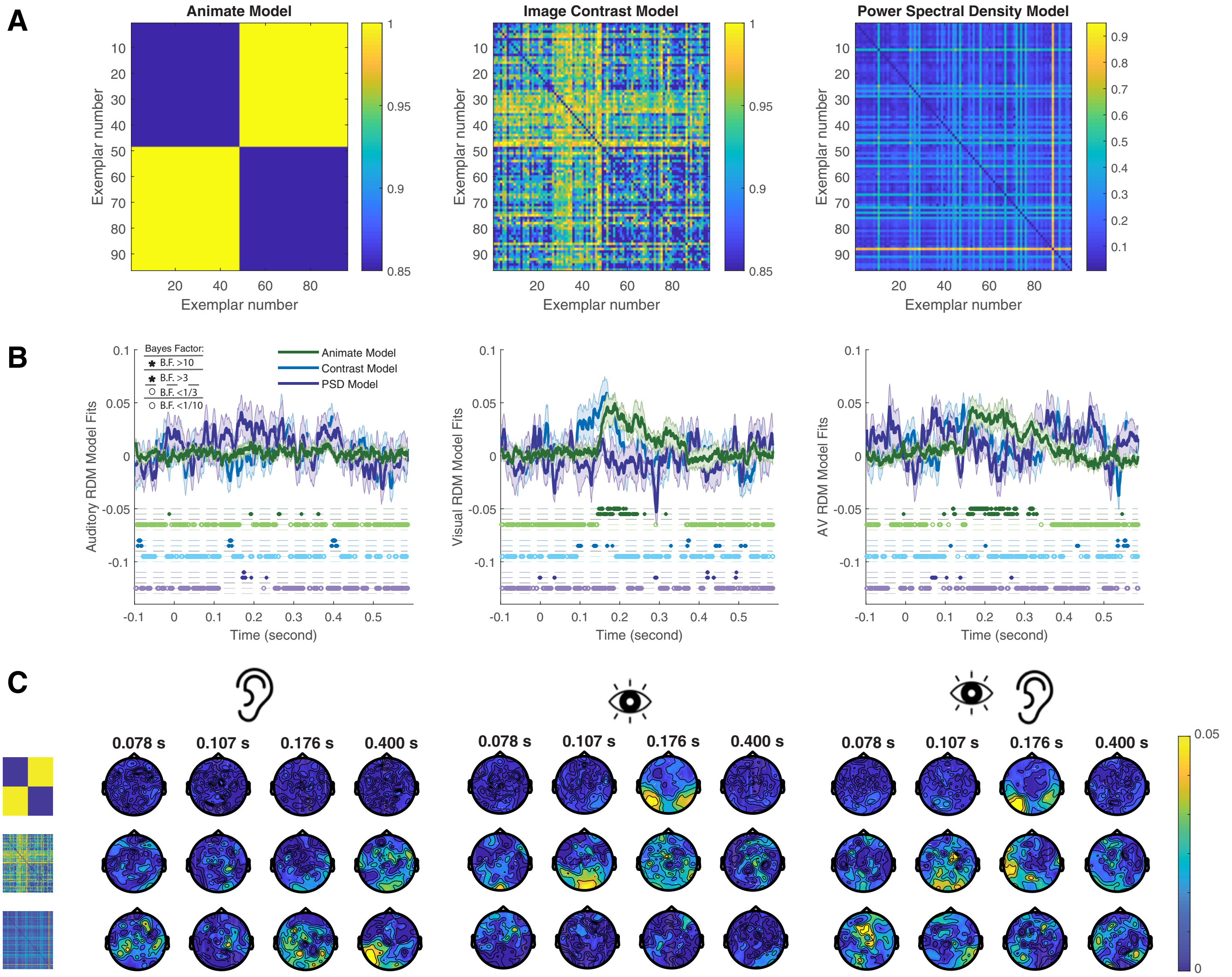

Figure 7. Model testing: abstract category models predict neural activity better than low-level feature models. ***A***, Category and low-level visual and auditory feature models. The animacy category model was constructed using a "0" for within-animacy category exemplars and "1" for between-animacy category exemplars. For the image contrast RDM, since all images were black and white drawings, the contrast vector consisted of the intensity values of each image. The power spectral density RDM was built using a Welch's power spectral density estimate and converted to a single vector for each sound. Each RDM was then constructed by taking the Spearman distance of each respective pairwise stimulus comparison. ***B***, Each model RDM was then tested with the auditory, visual, and audiovisual time-resolved RDMs on a subject-by-subject basis. Shaded area around lines indicate SE across subjects, with asterisks indicating substantial and strong evidence for the alternative hypothesis (BF $>$ 3 and $>$10) of a correlation $>$0 and null hypothesis (BF $<$ 1/3 and $<$1/10). ***C***, Model testing performed on electrode-specific RDMs using a searchlight analysis.

tested for significance using Bayes factors (BF $>$ 3). The contrast model and power spectrum model only had sporadic time points that had substantial evidence for the alternative hypothesis. The power spectrum model was most correlated with the auditory RDM with time points between 170 and 200 ms post-stimulus presentation, while the contrast model was most correlated with the visual RDM from 90 to 170 ms post-stimulus presentation. Further, as shown in Figure 7C, at early times, such as 107 ms, the occipital electrodes are most correlated with the contrast model. Similarly, for the auditory RDMs, temporoparietal electrodes correlate most with the power spectrum model early at 78 ms and late in the time course at 400 ms. In contrast, when we used an abstract model that ignored low-level features and instead separated stimuli based on object animacy category, we found a significant correlation (BF $>$ 3) with the visual RDMs beginning at 150 ms and audiovisual RDMs at 158 ms. Occipital and temporal electrodes for visual and audiovisual presentations were most correlated with the animacy model at timepoints such as 176 ms, but not later at 400 ms. The animacy model did not show a sustained correlation with the auditory RDM, implying that the animacy distinction is not as prominent in audition.

Discussion

In this study, we leveraged the visual and auditory encoding bias that has been observed for animate objects over inanimate objects (Murray et al., 2006; Vogler and Titchener, 2011; Tzovara et al., 2012; Guerrero and Calvillo, 2016; Grootswagers et al., 2017b) to study how perceptual biases across object categories influence the multisensory enhancement of audiovisual objects. Using behavioral measures and neural decoding, we found additional support for previous findings showing visual and auditory perceptual advantages for animate objects over inanimate objects. However, and somewhat surprisingly, we found that the advantage for animacy was not evident when objects were presented as audiovisual objects. Using RSA, we show that the lack of an animacy bias in audiovisual objects is in the context of an overall expansion of representational space when compared with visual and auditory objects. Further analysis showed that audiovisual presentations preferentially enhanced neural decoding of inanimate objects. A searchlight analysis and representational connectivity analysis showed that the presentation of inanimate objects in an audiovisual context may improve their encoding through increased representational connectivity between brain areas. We finally linked neural decoding and behavioral performance by using a distance-to-bound analysis and found that improved neural decoding for visual and audiovisual objects was associated with faster reaction times in the animacy categorization task. Furthermore, the decoding differences between visual and audiovisual objects were also predictive of their reaction time differences. Together, the results of our study provide new insights into the encoding of unisensory and multisensory objects, establish critical links between neural activity and behavior in the context of object categorization, as well as explore potential mechanistic differences in multisensory integration for weakly and strongly encoded objects.

Although stimulus features clearly contribute to the formation of object categories, including the distinction between animate and inanimate objects, there is ample evidence that the animate-inanimate distinction transcends stimulus features and can be thought of as an abstract category distinction. The distinction is present for stimuli presented in both the visual and auditory modalities, suggesting that animacy is a general organizing principle. Furthermore, category-specific deficits in naming animate objects have been found in patients who have experienced brain damage (Vignolo, 1982, 2006; Warrington and Mccarthy, 1987; Clarke et al., 2002; Kolinsky et al., 2002; Capitani et al., 2003). The category distinction is preserved across species; being present in both monkey inferotemporal (IT) cortex and human IT cortex. Furthermore, the use of carefully controlled stimuli that account for stimulus features have reinforced the categorical nature of animacy (Bracci et al., 2017; Ritchie and Op De Beeck, 2019). Similarly, auditory studies have also provided evidence for animacy as an abstract category distinction (Murray et al., 2006; De Lucia et al., 2010; Giordano et al., 2013). In the current study, we corroborate these findings by showing a significant correlation between an animacy model and neural response patterns, but a lack of consistent correlations between low-level stimulus features such as visual contrast and auditory power spectrum with neural response patterns.

Our study showed overall magnitude and temporal enhancement for audiovisual objects over visual and auditory objects consistent with recent findings (Brandman et al., 2019; Mercier and Cappe, 2019), and we additionally provide new insights into how audiovisual benefits selectively enhance the category of inanimate objects. Specifically, we found that the animacy bias for auditory and visual objects is absent in audiovisual objects. We hypothesized that the brain may be preferentially integrating the visual and auditory components of the more weakly encoded inanimate objects. Thus, greater multisensory integration for inanimate objects may serve to close the perceptual gap between animate and inanimate objects, consistent with the concept of inverse effectiveness (Stein and Meredith, 1993; Wallace et al., 2004). To test whether there were behavioral differences in multisensory integration across categories, we examined our behavioral data for a prediction made by maximum likelihood estimate models (Ernst and Banks, 2002), as follows: there is stronger multisensory benefit when the unisensory reliability or other measure of variability between senses is closer (i.e., smaller differences between visual and auditory reaction time). In agreement, we found that smaller RT differences between visual and auditory objects led to faster multisensory reaction times for inanimate objects, but not for animate objects. In the same vein, the neural decoding bias for animate over inanimate objects was no longer present for audiovisual presentations. When we subtracted audiovisual decoding from visual decoding, we found that decoding was only enhanced for inanimate objects, lending further evidence that audiovisual presentations selectively improved the encoding of inanimate objects.

To investigate the potential mechanism by which audiovisual presentations asymmetrically enhance the decoding of inanimate objects, we used representational connectivity analysis across all EEG sensors. Representational connectivity analysis has been previously used in a more limited way to assess representational similarity between two brain areas (Kriegeskorte et al., 2008b). In our analysis, we used a moving searchlight consisting of each electrode and its immediate surrounding neighbors to measure the different patterns of activity for each given stimulus. By doing so, we are able to use RCA as a tool to acquire a data-driven measure of how similar response patterns are topographically arranged across the brain. We predicted that animate and inanimate exemplars might demonstrate differences in connectivity measures, as previous studies have shown increased connectivity for biologically plausible motion over mechanical motion (Hillebrandt et al., 2014). Note that in this analysis, neighboring electrodes will have shared signals simply due to proximity. Therefore, the importance of these connectivity measures is the relative difference between animate and inanimate categories. We found increased representational connectivity for animate objects when presented in vision and when compared with inanimate objects. However, much like for our RSA results, these connectivity differences were no longer present for when these objects were presented in an audiovisual context. Additionally, the connectivity increase for inanimate objects occurs within the 100–200 ms time epoch that we have previously noted as the time period in which audiovisual presentations showed the greatest enhancement over visual presentations. One possible explanation for these results is that there may be increased audiovisual integration for inanimate objects relative to animate objects, leading to greater spread of neural representation across brain areas. However, the current analysis cannot exclude the possibility that the increase in inanimate connectivity for audiovisual presentations may also be due a more localized spread within electrodes in close proximity.

Next, we directly linked the neural results to behavioral results at the exemplar level by using a distance-to-bound approach (Carlson et al., 2014; Ritchie et al., 2015; Grootswagers et al., 2017b). This approach is a data-driven way of determining the relationship between neural representational space and behavioral measures (i.e., reaction times). In this analysis, we found a significant relationship between visual and audiovisual decoding distances and reaction times during two distinct poststimulus time epochs. One corresponded to peak decoding in our RSA analysis (i.e., 100–200 ms) and the other emerged ~150 to 200 ms later. These intervals and the corresponding topographic analyses in Figure 6*A* correspond to periods and electrodes associated with sensory evidence accumulation and decision-making, respectively (Murray et al., 2006; Tzovara et al., 2012). We next directly correlated multisensory neural decoding enhancements to reaction time improvements. Interestingly, we found that, despite an overall neural enhancement for audiovisual presentations, some exemplars showed possible effects of audiovisual interference effects. In these cases, visual decoding distances were greater than audiovisual decoding distances. These effects were largely reflected in the reaction time differences between audiovisual and visual presentation, with an overall significant negative correlation between behavioral audiovisual enhancement and neural audiovisual enhancement. These results provide evidence that the added sensory information in audiovisual presentations did not just provide the classifier with more information, but in fact provide further value for the object categorization task (Grootswagers et al., 2018). However, it does not eliminate the possibility that added neural information was also used for other aspects of the perceptual response not tapped in the current paradigm (e.g., response confidence).

In conclusion, our study introduces new insights into the representation by the brain of sensory and multisensory information as it relates to object encoding. The greater neural encoding benefits for inanimate stimuli seen under audiovisual conditions compliments prior work, where sensory information was selectively removed from object stimuli, resulting in a selective contraction of the representational space of animate objects (Grootswagers et al., 2017b). Collectively, these findings show that neural representational space and the encoding of objects are impacted by both semantic congruence and stimulus modality (stimulus combinations) in a dynamic fashion. Future directions of our current work include approaches to investigate the interplay between parametrically reducing neural encoding by degrading visual stimuli, while simultaneously using audiovisual presentations to enhance neural encoding. Understanding the computational framework the brain uses to maximize the sensory information it captures across sensory systems has broad implications for how stimuli perturbations and sensory integration affect object encoding.

References

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309–312.

Berger JO, Delampady M (1987) Testing precise hypotheses. Statist Sci 2:317–335.

Berry DA, Hochberg Y (1999) Bayesian perspectives on multiple comparisons. J Stat Plan Inference 82:215–227.

Bracci S, Ritchie JB, de Beeck HO (2017) On the partnership between neural representations of object categories and visual features in the ventral visual pathway. Neuropsychologia 105:153–164.

Braga RM, Hellyer PJ, Wise RJS, Leech R (2017) Auditory and visual connectivity gradients in frontoparietal cortex. Hum Brain Mapp 38:255–270.

Brandman T, Avancini C, Leticevscaia O, Peelen MV (2019) Auditory and semantic cues facilitate decoding of visual object category in MEG. Cereb Cortex 30:597–606.

Capitani E, Laiacona M, Mahon B, Caramazza A (2003) What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. Cogn Neuropsychol 20:213–261.

Cappe C, Thelen A, Romei V, Thut G, Murray MM (2012) Looming signals reveal synergistic principles of multisensory integration. J Neurosci 32:1171–1182.

Carlson T, Tovar DA, Alink A, Kriegeskorte N (2013) Representational dynamics of object vision: the first 1000 ms. J Vis 13(10):1, 1–19.

Carlson TA, Ritchie JB, Kriegeskorte N, Durvasula S, Ma J (2014) Reaction time for object categorization is predicted by representational distance. J Cogn Neurosci 26:132–142.

Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. Nat Neurosci 17:455–462.

Clarke S, Bellmann Thiran A, Maeder P, Adriani M, Vernet O, Regli L, Cuisenaire O, Thiran J-P (2002) What and where in human audition: selective deficits following focal hemispheric lesions. Exp Brain Res 147:8–15.

De Lucia M, Clarke S, Murray MM (2010) A temporal hierarchy for conspecific vocalization discrimination in humans. J Neurosci 30:11210–11221.

Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. Science 293:2470–2473.

Edwards W, Lindman H, Savage LJ (1963) Bayesian statistical inference for psychological research. Psychol Rev 70:193–242.

Ernst MO, Banks MS (2002) Humans Integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433.

Gelman A, Tuerlinckx F (2000) Type S error rates for classical and Bayesian single and multiple comparison procedures. Comput Stat 15:373–390.

Gelman A, Hill J, Yajima M (2012) Why we (usually) don't have to worry about multiple comparisons. J Res Educ Eff 5:189–211.

Giordano BL, McAdams S, Zatorre RJ, Kriegeskorte N, Belin P (2013) Abstract encoding of auditory objects in cortical activity patterns. Cereb Cortex 23:2025–2037.

Grootswagers T, Wardle SG, Carlson TA (2017a) Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. J Cogn Neurosci 29:677–697.

Grootswagers T, Ritchie JB, Wardle SG, Heathcote A, Carlson TA (2017b) Asymmetric compression of representational space for object animacy categorization under degraded viewing conditions. J Cogn Neurosci 29:1995–2010.

Grootswagers T, Cichy RM, Carlson TA (2018) Finding decodable information that can be read out in behaviour. Neuroimage 179:252–262.

Grootswagers T, Robinson AK, Carlson TA (2019) The representational dynamics of visual objects in rapid serial visual processing streams. Neuroimage 188:668–679.

Guerrero G, Calvillo DP (2016) Animacy increases second target reporting in a rapid serial visual presentation task. Psychon Bull Rev 23:1832–1838.

Hillebrandt H, Friston KJ, Blakemore SJ (2014) Effective connectivity during animacy perception—dynamic causal modelling of human connectome project data. Sci Rep 4:6240.

Jackson RE, Calvillo DP (2013) Evolutionary psychology. Evol Psychol 11:1011–1026.

Jarosz AF, Wiley J (2014) What are the odds? A practical guide to computing and reporting Bayes factors. J Probl Solving 7:2–9.

Jeffreys H (1998) The theory of probability. New York: Oxford UP.

Johnson VE (2013) Revised standards for statistical evidence. Proc Natl Acad Sci U S A 110:19313–19317.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Kolinsky R, Fery P, Messina D, Peretz I, Evinck S, Ventura P, Morais J (2002) The fur of the crocodile and the mooing sheep: a study of a patient with a category-specific impairment for biological things. Cogn Neuropsychol 19:301–342.

Kriegeskorte N, Mur M, Bandettini P (2008a) Representational similarity analysis—connecting the branches of systems neuroscience. Front Syst Neurosci 2:4.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008b) Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60:1126–1141.

Laws KR (2000) Category-specific naming errors in normal subjects: the influence of evolution and experience. Brain Lang 75:123–133.

Lindh D, Sligte IG, Assecondi S, Shapiro KL, Charest I (2019) Conscious perception of natural images is constrained by category-related visual features. Nat Commun 10:4106.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.

Mercier MR, Cappe C (2019) The interplay between multisensory integration and perceptual decision making. BioRxiv. Advance online publication. Retrieved Jan 7, 2019. doi: 10.1101/513630.

Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S (2006) Rapid brain discrimination of sounds of objects. J Neurosci 26:1293–1302.

Nastase SA, Connolly AC, Oosterhof NN, Halchenko YO, Guntupalli JS, Visconti di Oleggio Castello M, Gors J, Gobbini MI, Haxby JV (2017) Attention selectively reshapes the geometry of distributed semantic representation. Cereb Cortex 27:4277–4291.

New J, Cosmides L, Tooby J (2007) Category-specific attention for animals reflects ancestral priorities, not expertise. Proc Natl Acad Sci USA 104:16598–16603.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869.

Oosterhof NN, Connolly AC, Haxby JV (2016) CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in matlab/GNU octave. Front Neuroinform 10:27.

Ritchie JB, Carlson TA (2016) Neural decoding and "inner" psychophysics: a distance-to-bound approach for linking mind, brain, and behavior. Front Neurosci 10:190.

Ritchie JB, Op De Beeck H (2019) Using neural distance to predict reaction time for categorizing the animacy, shape, and abstract properties of objects. Sci Rep. 9:13201.

Ritchie JB, Tovar DA, Carlson TA (2015) Emerging object representations in the visual system predict reaction times for categorization. PLoS Comput Biol 11:e1004316.

Robinson AK, Grootswagers T, Carlson TA (2019) The influence of image masking on object representations during rapid serial visual presentation. Neuroimage 197:224–231.

Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G (2009) Bayesian t tests for accepting and rejecting the null hypothesis. Psychon Bull Rev 16:225–237.

Sellke T, Bayarri MJ, Berger JO (2001) Calibration of p values for testing precise null hypotheses. Am Stat 55:62–71.

Snodgrass JG, Vanderwart M (1980) A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J Exp Psychol 6:174–215.

Stein BE, Meredith MA (1993) The merging of the senses. Cambridge, MA: MIT.

Thelen A, Cappe C, Murray MM (2012) Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. NeuroImage 62:1478–1488.

Tzovara A, Murray MM, Plomp G, Herzog MH, Michel CM, De Lucia M (2012) Decoding stimulus-related information from single-trial EEG responses based on voltage topographies. Pattern Recognit 45:2109–2122.

Vignolo L (1982) Auditory agnosia. Philos Trans R Soc Lond B Biol Sci 298:49–57.

Vignolo LA (2006) Music agnosia and auditory agnosia. Ann N Y Acad Sci 999:50–57.

Vogler JN, Titchener K (2011) Cross-modal conflicts in object recognition: determining the influence of object category. Exp Brain Res 214:597–605.

Wallace MT, Ramachandran R, Stein BE (2004) A revised view of sensory cortical parcellation. Proc Natl Acad Sci U S A 101:2167–2172.

Warrington EK, Mccarthy RA (1987) Categories of knowledge: further fractionations and an attempted integration. Brain 110:1273–1296.

Wetzels R, Matzke D, Lee MD, Rouder JN, Iverson GJ, Wagenmakers EJ (2011) Statistical evidence in experimental psychology: an empirical comparison using 855 t tests. Perspect Psychol Sci 6:291–298.

Yuval-Greenberg S, Deouell LY (2009) The dog's meow: asymmetrical interaction in cross-modal object recognition. Exp Brain Res 193:603–614.